FEBS Letters 580 (2006) 2041–2045

Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks

Ashwini Patil[a,b], Haruki Nakamura[a,*]

[a] Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
[b] Department of Biology, Graduate School of Science, Osaka University, Japan

Received 6 January 2006; revised 24 February 2006; accepted 2 March 2006

Available online 10 March 2006

Edited by Takashi Gojobori

Abstract **We investigate the structural properties of hubs that enable them to interact with several partners in protein–protein interaction networks. We find that hubs have more observed and predicted disordered residues with fewer loops/coils, and more charged residues on the surface as compared to non-hubs. Smaller hubs have fewer disordered residues and more charged residues on the surface than larger hubs. We conclude that the global flexibility provided by disordered domains, and high surface charge are complementary factors that play a significant role in the binding ability of hubs.**

Keywords: Protein–protein interactions; Interaction networks; Hubs; Disordered regions; Surface charge

1. Introduction

Protein–protein interaction networks are characterized by the presence of a few highly connected nodes called hubs [1–3]. Hubs can interact with their partners either simultaneously to form a complex, or at different times and locations, connecting various biological modules in the network [4]. Given their ability to interact with multiple partners, it is not surprising that hubs play a central role in all biological processes of the cell by defining the properties of the interaction network [5]. However, the structural properties that give hubs the ability to interact with multiple partners, and differentiate them from non-hubs, are as yet unexplored.

Intuitively, flexibility or the ability to fold into an ensemble of conformations may be an important property required by hubs to bind multiple proteins. This flexibility can be manifested in two forms: (1) local flexibility in the form of a large number of loops/coils, in a folded globular protein, that take on the appropriate conformations required to bind multiple partners, or (2) global flexibility in the form of unrestricted movement allowed by the presence of one or more disordered or unfolded regions which have no tertiary structure with little or no secondary structure [6]. Loops differ from disordered regions in that they have a fixed tertiary structure for a given conformation of the protein [7].

Disordered regions have been previously found in several proteins associated with cancer and cell signaling [8,9]. Some of these include hubs, like the tumor suppressor Breast cancer type-1 susceptibility protein (BRCA1) with a large central disordered region that acts as a binding domain and a flexible linker [10], and the cyclin-dependent kinase (cdk) inhibitor $p21^{Waf1/Cip1/Sdi1}$ with a region that undergoes a disorder-to-order transition by folding into a stable conformation on binding its target [11]. However, Liu et al. have reported that proteins with 'loopy regions' or those with no secondary structure (NORS) had 'slightly more' interaction partners than non-NORS proteins [12]. Hence, it is not clear whether the flexibility that hubs need comes from loops in folded structures or extended conformations in disordered regions.

Not all hubs have large flexible regions that may help them bind different proteins. The degradation tag protein, Ubiquitin, is a small hub without a known disordered region as is the electron transport protein, Ferredoxin. But both have highly charged surfaces. Charged and polar residues on the surface of proteins at binding interfaces are known to contribute to binding specificity and complex stabilization, as well as promiscuous binding [13,14]. Hence high surface charge is likely to affect the binding ability of hubs.

In this study, we attempt to identify general structural tendencies in hubs that facilitate their binding to multiple proteins and find some distinctive characteristics that differentiate hubs from non-hubs. Specifically, we look for enrichment or depletion of disordered regions, loops and high surface charge in hubs as compared to non-hubs.

2. Materials and methods

2.1. Data set

All interactions were taken from the Database of Interacting Proteins (DIP) (July 2004) [15], and IntAct (March 2005) [16]. Interactions from high-throughput experiments were filtered for likely true positives [17]. Secondary structure assignments were calculated using DSSP [18]. Protein sequences were clustered at 50% sequence identity using CD-HIT [19]. Proteins with more than five interactions were considered as hubs, while those with one interaction are considered as non-hubs [1,4]. Non-hubs with two to five interactions were eliminated from the analyses to reduce the number of false positives. This resulted in a total of 1662 hubs and 4120 non-hubs with an average sequence identity of 26.5% (e-value = 10^{-8}) within each group. Of these, 222 hubs and 425 non-hubs had structures in Protein Data Bank (PDB) [20] (Supplementary Tables 1 and 2).

*Corresponding author. Fax: +81 6 6879 8636/4310.
E-mail address: harukin@protein.osaka-u.ac.jp (H. Nakamura).

Abbreviation: PDB, Protein Data Bank

0014-5793/$32.00
doi:10.1016/j.febslet.2006.03.003

2.2. Characteristics

Observed disordered residues. These were calculated for proteins with structures in PDB using: (1) residues with missing electron densities, indicated in 'REMARK 465' in the PDB files, from X-ray crystal structures having a resolution of at least 2.5 Å in regions having more than 10 consecutive residues, and (2) residues with no known structure to account for disordered regions removed during target selection or purification prior to structure determination due to their inability to fold.

Predicted disordered residues. Disordered residues were predicted for all protein sequences using DISOPRED2 with a false positive rate of 5% [21]. To further reduce the false positive rate, only those residues that were part of predicted disordered regions of length 30 or more were considered, since these are optimally recognized by DISOPRED2 [21].

Loops. Residues in loops were identified based on the secondary structure assignments of DSSP as: (1) those that were not assigned as α helix (H), 3_{10}-helix (G) or β strand (E) to account for rigid loops having limited flexibility, and (2) those with missing electron density in segments of less than 10 consecutive residues, in X-ray crystal structures having a resolution of at least 2.5 Å, to account for flexible loops or wobbly domains.

Fractions of disordered residues, and those in loops, were calculated over the length of the protein.

Charged residues on the surface. These included all the charged residues (Arg + Lys + Glu + Asp) in the protein structure with a solvent accessible area greater than 60 $Å^2$, as given by DSSP. We calculated the fraction of charged residues on the protein surface over the total number of residues on its surface.

Amino acid propensity. The propensity for each amino acid was calculated as the fraction of the amino acid in all the residues of 1662 hubs and 4120 non-hubs and the surface residues for 222 hubs and 425 non-hubs, to obtain the relative difference in propensity between hubs and non-hubs for each amino acid.

2.3. Statistical methods

We calculated the mean or the average for the test values of each characteristic listed above for hubs and non-hubs and used the difference in their means as our test statistic. Since the population distribution was not Gaussian, we calculated the 95% confidence intervals (CI) for the means using the standard error of means (SEM) [22]. We used randomization without replacement to calculate the significance of the difference in means, or the *P*-value [23] (Refer Supplementary Materials).

3. Results

3.1. Hubs have more disordered residues

Fig. 1 shows the differences in the means of several characteristics of hubs and non-hubs. We find that on an average, hubs have 28.32% more observed disordered residues ($p = 0.0022$). Hubs also have a higher percentage of predicted disordered residues at 26.29% ($p = 0.0002$). Surprisingly, hubs have 14.30% fewer residues in loops than non-hubs ($p = 0.0046$). This implies that the disordered regions have a greater impact than loops on the binding ability of hubs.

The difference in the amino acid propensities of hubs and non-hubs based on their amino acid sequences is shown in Fig. 2. Hubs have more charged and polar residues while non-hubs have more hydrophobic residues. Residues like Glu, Lys, Ser, Gln, which are commonly found in disordered regions of proteins, are enriched in hubs. On the other hand, Trp, Cys, Phe, which are commonly found in ordered regions, are enriched in non-hubs [24]. Disordered regions are known to be characterized by a high net charge and a low mean hydrophobicity [25]. Hence, we conclude that hubs have more disordered residues than non-hubs.

However, we do not find any correlation between the number of interactions and the percentage of predicted or observed disordered residues in hubs (Pearson's correlation coefficient, $r = 0.020$ and $r = -0.067$, respectively).

3.2. Hubs have more charged residues on the surface

As shown in Fig. 1, we find that hubs have 6.81% ($p = 0.0018$) more charged residues on the surface than non-hubs. In Fig. 2, the difference in the amino acid propensities of the surface residues of hubs and non-hubs shows hubs enriched in charged residues as compared to non-hubs. These observations implicate high surface charge as a factor affecting the promiscuity of hubs.

We also find that hubs with fewer observed disordered residues have more charged residues on the surface. Hubs with less than 10% disordered residues have an average of 49 charged residues on the surface as compared to the average 42 charged residues on the surfaces of hubs with more than 10% disordered residues (see Fig. 3A). This suggests that in the absence of disordered regions, a higher surface charge may help hubs bind multiple partners. Not surprisingly, most of the hubs with less than 10% disordered residues are small proteins. Almost 50% of the hubs with less than 10% disordered residues are shorter than 250 residues as compared to 16% of those with

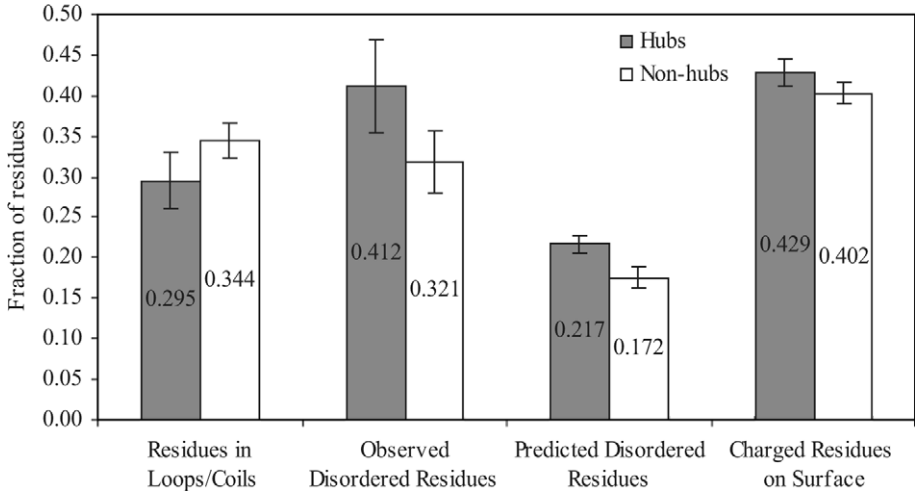

Fig. 1. Mean percentage of residues found in loops/coils, observed disordered residues, predicted disordered residues and charged surface residues in structures of hubs (gray) and non-hubs (white). Error bars denote the upper and lower 95% confidence intervals of the means.

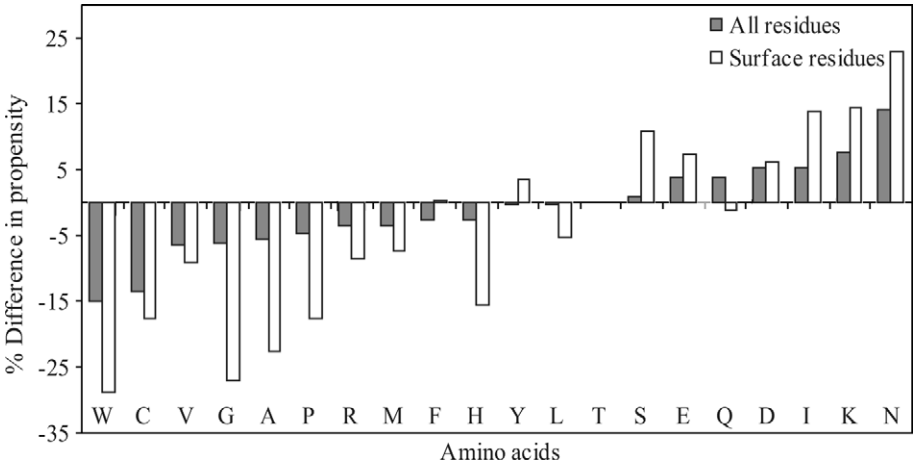

Fig. 2. Difference in amino acid propensities of all residues (gray) and surface residues (white) of hubs and non-hubs with positive values showing enrichment of amino acids in hubs and negative values showing enrichment of those in non-hubs.

more than 10% disordered residues (see Fig. 3B). This shows that in small hubs, where disordered residues tend to be few or absent, higher surface charge plays a more dominant role in the binding ability.

We conclude that disorder and surface charge are complementary characteristics that help hubs form multiple interactions. The effect of disordered domains is the dominant factor, especially in large hubs. In small hubs with few or no disordered domains, the effect of high surface charge is dominant.

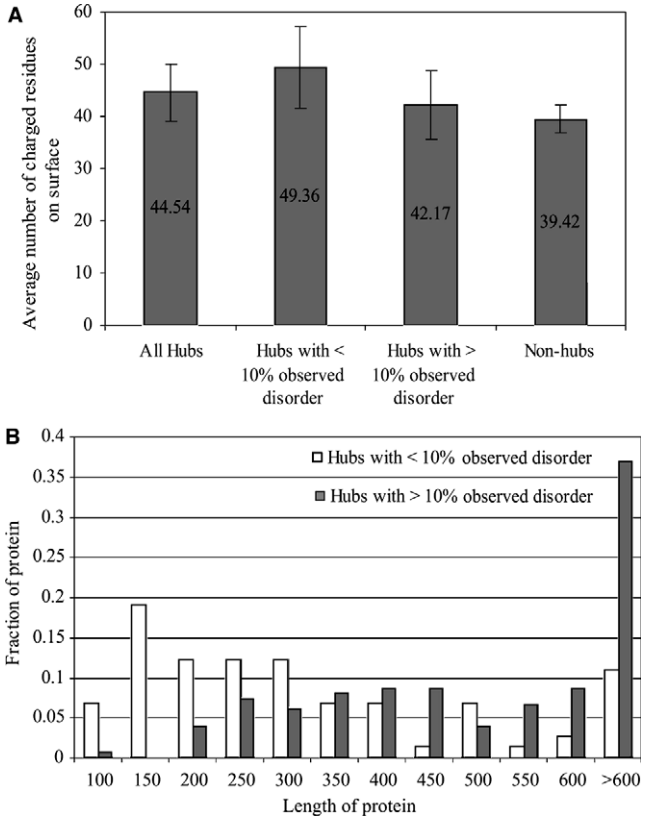

Fig. 3. (A) Average number of charged residues on the surface of hubs and non-hubs. Error bars denote the upper and lower 95% confidence intervals of the means. (B) Distribution of different categories of hubs with respect to their length.

4. Discussion

Hubs control the properties of a protein–protein interaction network through their ability to interact with multiple partners. We identify the structural properties of hubs that facilitate this.

We find that hubs have more disordered residues than non-hubs. This is further corroborated by the observation that hubs have a higher propensity for disorder promoting residues like Glu, Lys, Ser, Gln and a lower propensity for order promoting residues like Trp, Cys, Phe, Leu. Surprisingly, we do not find Proline as prevalent in hubs as expected [26]. The ability of Proline to prevent the formation of secondary structures may hinder the disorder-to-order transition that several hubs undergo on binding their target proteins resulting in its low propensity. The global flexibility provided by disordered regions, or domains, in hubs appears to have a greater impact on their binding characteristics than the local flexibility in the form of loops in folded structures.

The presence of disordered domains can provide various advantages to hubs, including global flexibility and induced folding [27], increased speed of interaction [28], and tight regulation through rapid turnover [6]. The disordered domain in a hub may be the binding site with different induced folding states depending on the target as seen in the N-terminal disordered domain of the transcription factor p53 [29], or a flexible linker that joins two ordered globular domains allowing them to move freely with respect to each other, as illustrated by the central disordered domain of Ca^{2+} bound Calmodulin [30] (see Fig. 4). Indeed, several hubs with disordered domains have been discussed by Dunker et al. [31].

We do not see any correlation between the fraction of disordered residues and the number of interactions in hubs. It is possible that there is no correlation between them. A protein with one or more disordered domains may have the ability to bind multiple partners but the number of its interaction partners may depend on other factors like cellular localization and expression levels. The other reason for this lack of correlation could be that our data set is not exhaustive in terms of

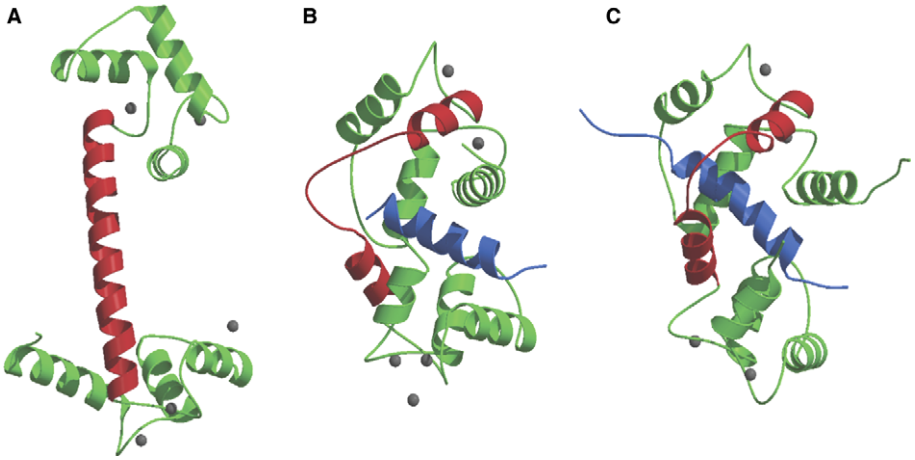

Fig. 4. (A) Ca^{2+}–Calmodulin complex (1EXR). (B) Ca^{2+}–Calmodulin complex bound to calmodulin-dependent protein kinase II-alpha (blue) (1CM1). (C) Ca^{2+}–Calmodulin complex bound to myosin light chain kinase (blue) (2BBM). Ca^{2+} binding domains of calmodulin are shown in green, flexible linker in red and Ca^{2+} in gray.

interactions or disordered domains. The observed disordered residues may include those that do not have a structure in PDB because they are not of considerable interest or significance, while missing those in complex structures that have acquired a stable conformation on binding a target. Also, limiting the predicted disordered residues to those in regions of at least 30 consecutive residues eliminates shorter disordered regions.

Apart from disordered domains, we show that high surface charge is the other important characteristic of hubs that is likely to have an impact on their binding ability. We also show that disorder and high surface charge are complementary factors in hubs, with high surface charge having a dominant effect on the binding ability in the absence of disordered domains. In the case of small hubs with few or no disordered domains, surfaces may be characterized by very high charges to allow multiple interactions (Supplementary Fig. 1).

Undoubtedly, structural properties are not the only ones to play a defining role in the ability of a hub to interact with multiple partners. Other important properties such as the localization of the protein and its possible targets in the cell, the regulation of their expression and degradation, and the binding affinity of the different targets will also have a large impact on the number of proteins it interacts with.

Acknowledgement: This study was supported by Grant-in-Aid for Scientific Research on priority areas No. 17017024 from the Ministry of Education, Science, Sports and Culture of Japan.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2006.03.003.

References

[1] Jeong, H., Mason, S.P., Barabasi, A.L. and Oltvai, Z.N. (2001) Lethality and centrality in protein networks. Nature 411, 41–42.
[2] Giot, L. et al. (2003) A protein interaction map of *Drosophila melanogaster*. Science 302, 1727–1736.
[3] Li, S. et al. (2004) A map of the interactome network of the metazoan *C. elegans*. Science 303, 540–543.
[4] Han, J.-D.J. et al. (2004) Evidence for dynamically organized modularity in the yeast protein–protein interaction network. Nature 430, 88–93.
[5] Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. Nat. Rev. Genet. 5, 101–113.
[6] Wright, P.E. and Dyson, H.J. (1999) Intrinsically unstructured proteins: reassessing the protein structure–function paradigm. J. Mol. Biol. 293, 321–331.
[7] Bracken, C., Iakoucheva, L.M., Romero, P.R. and Dunker, A.K. (2004) Combining prediction, computation and experiment for the characterization of protein disorder. Curr. Opin. Struct. Biol. 14, 570–576.
[8] Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z. and Dunker, A.K. (2004) The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 32, 1037–1049.
[9] Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradovic, Z. and Dunker, A.K. (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. J. Mol. Biol. 323, 573–584.
[10] Mark, W.Y., Liao, J.C., Lu, Y., Ayed, A., Laister, R., Szymczyna, B., Chakrabartty, A. and Arrowsmith, C.H. (2005) Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein–protein and protein–DNA interactions. J. Mol. Biol. 345, 275–287.
[11] Kriwacki, R.W., Hengst, L., Tennant, L., Reed, S.I. and Wright, P.E. (1996) Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc. Natl. Acad. Sci. USA 93, 11504–11509.
[12] Liu, J., Tan, H. and Rost, B. (2002) Loopy proteins appear conserved in evolution. J. Mol. Biol. 322, 53–64.
[13] Friedler, A., Veprintsev, D.B., Rutherford, T., von Glos, K.I. and Fersht, A.R. (2005) Binding of Rad51 and other peptide sequences to a promiscuous, highly electrostatic binding site in p53. J. Biol. Chem. 280, 8051–8059.
[14] Sheinerman, F.B., Norel, R. and Honig, B. (2000) Electrostatic aspects of protein–protein interactions. Curr. Opin. Struct. Biol. 10, 153–159.
[15] Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U. and Eisenberg, D. (2004) The database of interacting proteins: 2004 update. Nucleic Acids Res. 32, D449–D451.
[16] Hermjakob, H. et al. (2004) IntAct: an open source molecular interaction database. Nucleic Acids Res. 32, D452–D455.
[17] Patil, A. and Nakamura, H. (2005) Filtering high-throughput protein–protein interaction data using a combination of genomic features. BMC Bioinformatics 6, 100.

[18] Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637.
[19] Li, W., Jaroszewski, L. and Godzik, A. (2002) Tolerating some redundancy significantly speeds up clustering of large protein databases. Bioinformatics 18, 77–82.
[20] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The protein data bank. Nucleic Acids Res. 28, 235–242.
[21] Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. and Jones, D.T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J. Mol. Biol. 337, 635–645.
[22] Motulsky, H. (1995) Intuitive Biostatistics, Oxford University Press, New York.
[23] Manly, B.F.J. (1997) Texts in Statistical Science, Chapman & Hall, London, New York.
[24] Dunker, A.K. et al. (2001) Intrinsically disordered protein. J. Mol. Graph. Model. 19, 26–59.
[25] Uversky, V.N., Gillespie, J.R. and Fink, A.L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41, 415–427.
[26] Lise, S. and Jones, D.T. (2005) Sequence patterns associated with disordered regions in proteins. Proteins 58, 144–150.
[27] Dyson, H.J. and Wright, P.E. (2005) Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell. Biol. 6, 197–208.
[28] Shoemaker, B.A., Portman, J.J. and Wolynes, P.G. (2000) Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proc. Natl. Acad. Sci. USA 97, 8868–8873.
[29] Dawson, R., Muller, L., Dehner, A., Klein, C., Kessler, H. and Buchner, J. (2003) The N-terminal domain of p53 is natively unfolded. J. Mol. Biol. 332, 1131–1141.
[30] Wilson, M.A. and Brunger, A.T. (2000) The 1.0 A crystal structure of Ca(2+)-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. J. Mol. Biol. 301, 1237–1256.
[31] Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M. and Uversky, V.N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J. 272, 5129–5148.